The Role of Weight on Community Structure of Networks

Ying Fan[1], Menghui Li[1], Peng Zhang[1], Jinshan Wu[2]*, Zengru Di[1]

1. Department of Systems Science, School of Management, Beijing Normal University, Beijing 100875, P.R.China
2. Department of Physics & Astronomy, University of British Columbia, Vancouver, B.C. Canada, V6T 1Z1

February 2, 2008

Abstract

The role of weight on the weighted networks is investigated by studying the effect of weight on community structures. We use weighted modularity Q^w to evaluate the partitions and Weighted Extremal Optimization algorithm to detect communities. Starting from idealized and empirical weighted networks, the distribution or matching between weights and edges are disturbed. Using dissimilarity function D to distinguish the difference between community structures, it is found that the redistribution of weights does strongly affect the community structure especially in dense networks. This indicates that the community structure in networks is a suitable property to reflect the role of weight.

Keyword: Weight, Community Structure, Dissimilarity

PACS: 89.75.Hc 05.40.-a 87.23.Kg

1 Introduction

Link Weights, as strength of the interaction represented by networks, are believed to be an important variable in networks. It gives more information about networks besides its topology properties dominated by links. Recently more and more study in complex networks focus on the weighted networks. The problems involve the definition of weight and other quantities which characterize the weighted networks[1, 2, 3], the empirical studies of its statistical properties[4, 5, 6, 7], evolving models[8, 9, 10, 11, 12, 13], and transportation or other dynamics on weighted networks[14, 15, 16, 17].

However, how important is the weight, or what significant changes on network structures are induced when weight is changed? This, we call, the role of weight, should be a fundamental question in the study of weighted networks. But it has not been investigated deeply in the previous studies.

*Author for correspondence: jinshanw@phas.ubc.ca

The role of weight should be first investigated by analyzing the correlation between edge weight and other properties. In this way, it attempt to answer the question that if there is some internal mechanism strongly determining weights. For example, one may image the edge betweenness effects edge weight largely because the larger edge betweenness implies the edge has the more important role in communication on networks so that the weight on the edge might be also larger. If this is true, the weight should be less important for the networks because the weight is dependent on links or the weight is determined by the network topology.

For the database of our scientific collaboration networks of econophysicists, BNU-email network, and monkey's societies, based on the standard method for linear correlation analysis in mathematical statistics, we get the correlation results for the edge weights and edge betweenness, the edge weight and the sum of degrees of its two ends vertices. The results are shown in Table 1. All the coefficients of correlation for links are less than 0.25. These negative results reveal that the weight is really an independent variable for complex networks. By the way, we also get the correlation coefficients for the vertex weight (strength of vertex) and its degree. They are 0.79, 0.44. and 0.71 for our above three networks. The results are rational because they are all contributed by the connected links.

Table 1. Correlation coefficients for weight and other quantities

		EP-SCN	BNU-Email	Monkey
Links	Weight-Betweenness	0.0055	0.028	0.19
	Weight-Degree of ends	0.226	0.028	0.14
Vertices	Strength-Degree	0.79	0.44	0.71

From the above negative conclusion on correlation analysis, we know the weight is an independent variable at some level. This even makes the work on the role of weight more attractive: since it's somehow independent, how significant is it?

The effects of weight on the network structures can be investigated on two classes of properties: single vertex statistics and correlation statistics. The former includes vertex based properties such as degree, clustering coefficient, and the later includes global properties such as distance, betweenness and especially community structure. Except we consider the difference of above properties between unweighted and weighted networks, an important way to study the effects of weight is to consider their difference after we disturb the weight distribution. A natural way to disturb weights is to redistribute weights onto different edges[7], where we either change the distribution of weight or mix the matching between weights and edges. However, our previous investigation shows that this redistribution has little effect on the single vertex statistics, neither significant effect on distance. This negative conclusion looks like weight does not have significant effect on network structures. However, we strongly suspect that it's only because we have not found the proper measurement to present its effect.

An analogy between networks and condensed matter may give us some clues for insightful investigation. In condensed matter, at most cases, an effective single electron picture is well enough for a large number of phenomena. An effective field is used to represent effect from all electrons and lattice ions in condensed matter. However, there are something beyond this single particle scheme so that it requires to consider the correlation between electrons. Single

vertex statistics naturally belong to the former class, even possible including some global properties such as distance. But community structure measures directly the correlation among vertices. Therefore, in this work, we use the community structure as a measurement on the role of weight.

In binary networks, the community structure is defined as groups of network vertices, within groups there are dense internal links among nodes, but between groups nodes loosely connected to the rest of the network[18]. However, as we have mentioned above, link weight is an independent variable and should have some important effects on structure and function of networks. As for community structures, the definition of the community must integrate links with link weights. Newman has generalized the modularity Q to weighted modularity Q^w for evaluation the partitions of weighted networks[2]:

$$Q^w = \frac{1}{2w}\sum_{ij}[w_{ij} - \frac{w_i w_j}{2w}]\delta(c_i, c_j), \tag{1}$$

It takes both links and link weights into account. Usually, groups separated with the link weights should be different from the result based only on topological linkage. Given the same topological structure, different assignments of link weights may result in different community structures. Our basic question is how will the community structure change when the weights are disturbed.

There are several questions should be answered before the realization of the above ideas. First, what are the networks for this investigation. Our previous analysis uses networks of Econophysicists as our typical networks[3]. Recently, we have got more datas on BNU-email networks and monkey's societies[2, 19]. Hopefully, dense weighted networks will give us more confirmative conclusions. Besides these real networks, we can also construct idealized ad hoc weighted networks for our investigation. Second, how to extract and to evaluate community structure from a given network. Currently, there are several typical algorithms in the literature: Hierarchical Clustering, betweenness based GN algorithm[20], Potts model based algorithm[24], Extremal Optimization algorithm[21], and so on. We have generalized several approaches to weighted networks and investigated their performance[22]. Here, we use only Weighted Extremal Optimization (WEO) algorithm. The approach of WEO is directly related with the definition of weighted modularity Q^w. It performs well in weighted networks. Then the third, how to compare different community structures among the same set of vertices. We have proposed dissimilarity function D in [19] to measure the difference between partitions. Starting from two community structures $\{A_1, A_2, \ldots\}$ and $\{B_1, B_2, \ldots\}$ over the same set N, first, we need to identify the correspondence between As and Bs and re-order them. Then second, for each pair after re-ordered, the dissimilarity of A_j and B_j is given by:

$$d_j = \frac{\left|\left(A_j \cap \bar{B}_j\right) \cup \left(\bar{A}_j \cap B_j\right)\right|}{|A_j \cup B_j|}. \tag{2}$$

And the total dissimilarity can be calculated as

$$D = \frac{\sum_{i=1}^{K} d_i}{K}. \tag{3}$$

Here, we use the dissimilarity function D to quantify the difference of different communities.

The paper is organized as following. The correlation between edge weight and betweenness, and other properties has been studied in Section 1. The negative results demonstrate that the weight is an independent variable. Then in Section 2, we compared community structures of weighted and corresponding unweighted, disturbed weighted networks. The results demonstrate that the weight has effects on communities, especially in dense networks. In Section 3, in order to show more results about the effects of weight on communities, we investigated idealized ad hoc weighted networks in detail. The results give us systematic view about the effects of weight on community structures. Finally we give some conclusion remarks.

2 Community structures in Real and Unweighted, Inverse Weighted Networks

In this section, we focus on the effect of weight on community structure in real weighted networks by comparing the communities of real and its binary correspondence, and also inverse weighted networks. Real networks include Econophysicists collaboration network[3], BNU-email network and Rhesus monkey network[2]. For detecting and comparing community structure, we take the largest connected cluster of the above networks. For the Econophysicists collaboration network, it includes 271 nodes and 371 edges. In order to distinguish the network with different proportion of possible links, we define the denseness of network as the ratio of existing links to the all possible links among the nodes. The denseness for Econophysicists collaboration network is 0.01. The database for BNU-Email network includes the times of Emails between any two mailboxes (*@bnu.edu.cn) in a week. The network includes 740 nodes and 1400 links. We also use its largest cluster, which includes 620 nodes and 1117 links. The denseness of BNU-Email network is 0.006. The monkey network include 16 nodes and 69 edges. It is a connected network with denseness equals 0.575. So it is a relatively dense network.

As mentioned in the introduction, besides considering the difference of suitable properties between unweighted and weighted networks, an important way to investigate the effects of weight is to study the impaction of weight redistribution to the network properties. We have introduced the way to re-assign weights onto edges with $p = 1, -1$ for weighted networks [7]. Set $p = 1$ represents the original weighted network given by the ordered series of weights which gives the relation between weight and edge but in a decreasing order,

$$W(p=1) = \left(w_{i_1 j_1} = w^1 \geq w_{i_2 j_2} = w^2 \geq \cdots \geq w_{(i_L)(j_L)} = w^L\right). \tag{4}$$

$p = -1$ is defined as the inverse order as

$$W(p=-1) = \left(w_{i_1 j_1} = w^L \leq \cdots \leq w_{(i_{L-1})(j_{L-1})} = w^2 \leq w_{(i_L)(j_L)} = w^1\right), \tag{5}$$

Here we only compare the community structures of original and inverse weighted networks.

We apply WEO algorithm 20 times for each network. Then the community structure is shown by the corresponding co-appearance matrix. Matrix ordering taken from a realization of WEO algorithm in original network. In order to show the difference between groups of original and unweighted, inverse weighted networks, we keep the same order as the original one in the co-appearance matrix of the unweighted and inverse weighted networks. Each

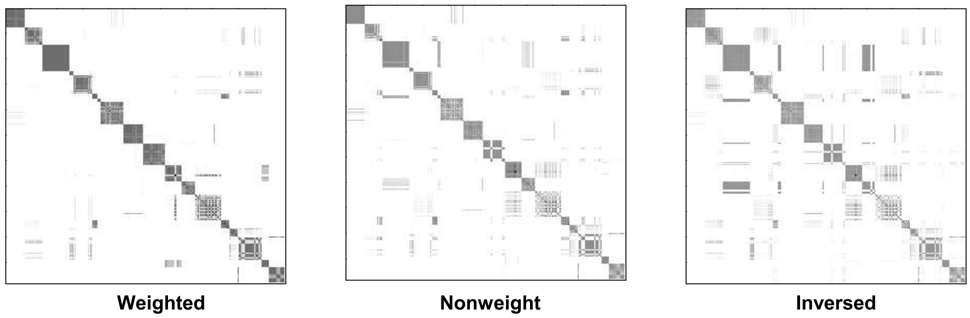

Figure 1: The normalized and colored co-appearance matrix for (a) original and (b) corresponding binary, and (c) inverse weighted Econophysicists collaboration networks.

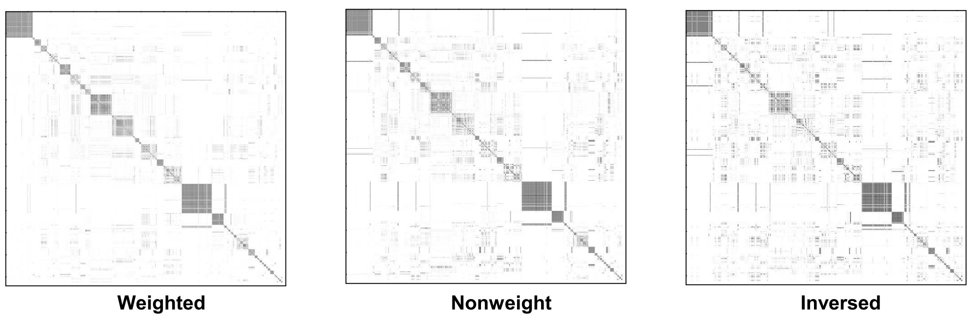

Figure 2: The normalized and colored co-appearance matrix for (a) original and (b) corresponding binary, (c) inverse weighted BNU-email networks.

matrix gives the fraction of nodes classified in the same partition over 20 realizations of WEO algorithm. The color of the position (i, j) corresponds to the fraction of times that nodes i and j belong to the same group. Then the final communities are given by the most probable partitions. The difference between any pair of partitions is given by the dissimilarity function D. In table 2, we show the comparison of the communities which formed in original and unweighted and inverse weighted networks (with $p = 1, -1$ respectively).

Table 2. Dissimilarities between communities of original and binary, inverse weighted networks

	EP-SCN	BNU-Email	Monkey
D(Original-Binary)	0.37	0.58	0.67
D(Original-Inverse)	0.46	0.69	0.75

As shown in the figures and Table 2, link weights indeed affect the community structure which is related to the global structure of networks. There are dissimilarities between

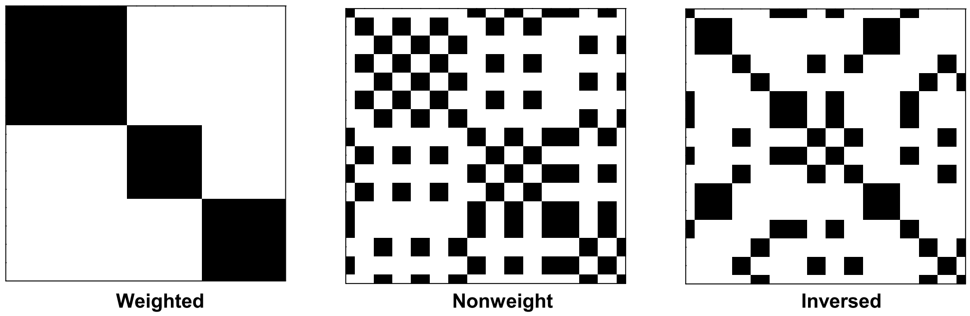

Figure 3: The normalized and colored co-appearance matrix for (a) original and (b) corresponding binary, (c) inverse weighted monkey networks. The precision of the algorithm for original and binary, inverse weighted networks are 1.00, 0.80, and 0.83 respectively. The dissimilarity of the results are much larger than the other two networks. It seems that weight plays more important role in dense networks.

partitions of weighted and its binary correspondence networks, especially for Rhesus monkey network. It seems that weight has more important effects on dense networks.

The WEO algorithm can only give us best partition of the network. Using GN algorithm, we can compare the process of dividing networks in to groups. Fig. 4 shows the dissimilarity between the results of original and binary, inverse weighted networks. It could be found that the dissimilarity usually has the larger value when the division of networks has gone into the inner structure of the community. That also indicates the weight has bigger influence in denser networks.

3 Results on Idealized Weighted Networks

Inspired by the above empirical studies, we constructed more idealized *ad hoc* weighted networks and try to get more systematic conclusions about the effects of weight.

The idealized networks is firstly introduced by Newman and used by many other authors[26, 23]. Each network consists of $n = 128$ vertices divided into four groups of 32 nodes. Vertices are assigned to groups and are randomly connected to vertices of the same group by an average of $\langle k_{intra} \rangle$ links and to vertices of different groups by an average of $\langle k_{inter} \rangle$ links. The average degree of all vertices are fixed, namely $\langle k_{intra} \rangle + \langle k_{inter} \rangle = 16$. With $\langle k_{intra} \rangle$ increasing from small, the communities become more and more diffuse, and it becomes more and more difficult to detect the communities. For a given network topology, here we assign similarity weight to each link. The intragroup link weight is assigned as w_{intra}, while the intergroup link weight is assigned as w_{inter}. In practise, the relationship among the nodes in groups is usually much closer than the relationship between groups. So w_{inter} is normally less than w_{intra}. Similarly with $\langle k_{intra} \rangle + \langle k_{inter} \rangle = 16$, we require

$$\langle w_{intra} \rangle + \langle w_{inter} \rangle = 2. \tag{6}$$

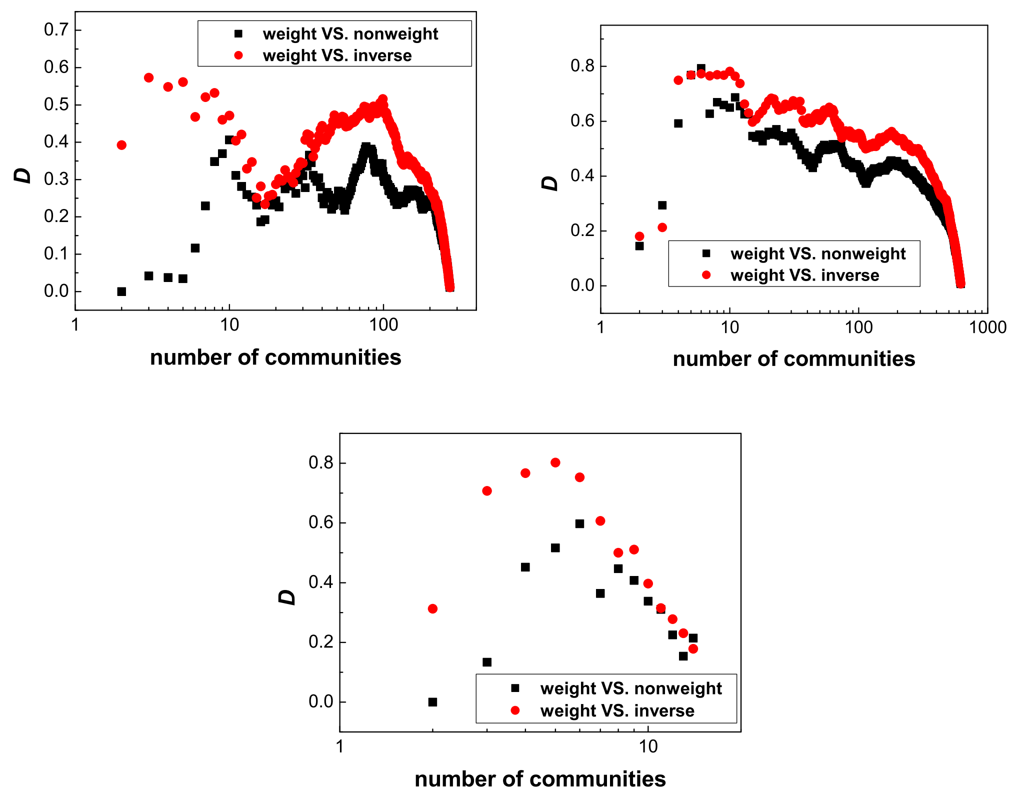

Figure 4: A: The dissimilarity of communities gotten by GN algorithm for original and binary, inverse weighted networks of (a)Econophysicists collaboration network, (b)BNU-email network, and (c)Monkey network.

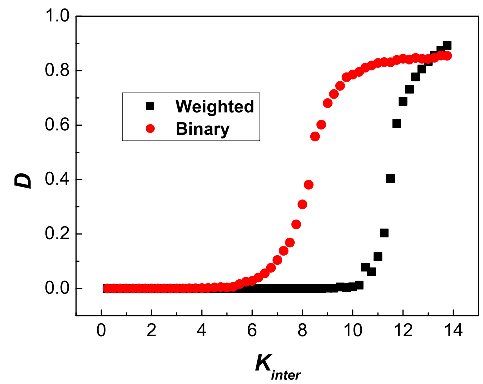

Figure 5: The dissimilarity between groups found by WEO algorithm and the presumed community structure. With the link weight assigned by the method described in the text, weighted networks have larger spectrum of $\langle k_{inter} \rangle$ than binary networks.

We use *ad hoc* networks with uniform distribution of link weights here. For a given network topology with certain $\langle k_{inter} \rangle$, weights are taken randomly from a 0.5 interval around $\langle w_{intra} \rangle$ and $\langle w_{inter} \rangle$ respectively for intragroup connections and intergroup connections, that is $\langle w_{intra} \rangle - 0.25$, $\langle w_{intra} \rangle + 0.25$ and $\langle w_{inter} \rangle - 0.25$, $\langle w_{inter} \rangle + 0.25$ respectively. In the following simulations, we take $\langle w_{intra} \rangle$=1.6 and so that $\langle w_{inter} \rangle$=0.4.

Now we exam the effects of weight on community structures based on idealized weighted networks. First, we applied WEO method to the binary and weighted *ad hoc* networks. The groups gotten by the algorithm are compared with the presumed communities with the dissimilarity function D. We could found in Fig.5, with the increasing of $\langle k_{inter} \rangle$, D increase sharply when $\langle k_{inter} \rangle$ is larger than 8 in binary networks. But in weighted networks, the spectrum of $\langle k_{inter} \rangle$ is much larger. Even when $\langle k_{inter} \rangle$ is around 10, The network could grouped correctly. For weighted networks, the community structure defined by weighted modularity Q^w integrate links with link weights. It shows that the weight has an important effect on community structure.

Then we compared the community structure of weighted networks with the groups after we inverse redistribute the link weight. The community structures are all gotten by WEO approach. The original idealized network is constructed with a given $\langle k_{inter} \rangle$ and $\langle w_{intra} \rangle$=1.5. Then we redistribute the link weights inversely according to the method described in Section 2. Fig.6 shows the dissimilarity function D between the original groups and groups after disturbing link weights. The results are the average of 20 network realizations and 10 runs each. It is interesting to find that D is around 0 when $\langle k_{inter} \rangle$ is small but it increases gradually with the increasing of $\langle k_{inter} \rangle$. When $\langle k_{inter} \rangle$ is large enough, it almost reaches 1. These results reveal that link and link weight are two factors that determine the structure of networks. When topological linkage dominates the structure of networks, link weight plays less important role in networks. On the other hand, in some networks, especially in dense networks such as Rhesus monkey network, link weight is crucial to the network structures.

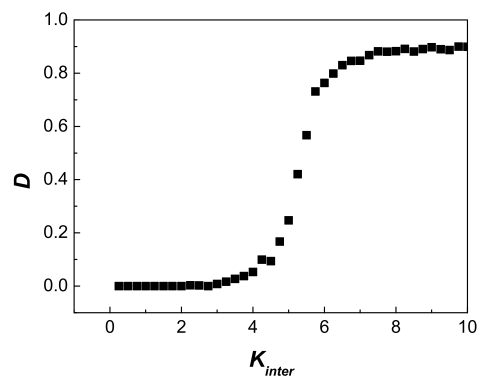

Figure 6: The dissimilarity between groups in original networks and the groups with random redistributed link weights. The communities are found by WEO algorithm. With the increase of $\langle k_{inter} \rangle$, link weights have more effects on community structures.

4 Concluding Remarks

In this paper, in order to investigate the role of weight, we pay much attention to the influence of the weight to the results of community structures. Besides the idealized ad hoc weighted networks, the econophysicist collaboration network, Rhesus monkey network, and BNU-email network are analyzed by using the WEO algorithm. Using the dissimilarity function D to measure the difference of two kinds of community structures, we investigate the different results of partition for non-weighted, weighted, and inverse weighted networks. It is found that weight do have influence on the formation of communities structure. That means: 1, the weight do have important role to the network structures; 2, the community structure is a suitable global properties to reflect the effect of weight. It has been also found that the weight is more significant for dense networks. For a sparse network, the existence or not of edges have bigger influence to community structure of networks than the weight.

References

[1] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Physica A, **346**(2005) 34-43.

[2] M. E. J. Newman, Phys. Rev. E. **70**(2004) 056131.

[3] Y. Fan, M. Li, J. Chen, L. Gao, Z. Di, J. Wu, International Journal of Modern Physics B, vol **18**(17-19)(2004) 2505-2511.

[4] E. Almaas, P. L. Krapivsky, and S. Redner, Phys. Rev. E **71**(2005) 036124.

[5] G. Bianconi, arXiv:cond-mat/0412399.

[6] W. Jezewski, Physica A **354**(2005) 672-680.

[7] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, and J. Wu, Physica A **350**(2005) 643-656.

[8] S. H. Yook, H. Jeong, A. Barabasi, and Tu. Y, Phys. Rev. Lett., **86**(2001) 5835C5838.

[9] A. Barrat, M. Barthelemy, and A. Vespignani, Phys. Rev. Lett., **92**(2004) 227801.

[10] T. Antal and P. L. Krapivsky, Phys. Rev. E, **71**(2005) (026103).

[11] A. Barrat, M. Barthelemy, and A. Vespignani, Phys. Rev. E, **70**(2004) (066149).

[12] S. N. Dorogovtsev and J. F. F. Mendes, arXiv:cond-mat/0408343.

[13] W.-X. Wang, B.-H. Wang, B. Hu, G. Yan, Q. Ou, Phys. Rev. Lett. **94**(2005) 188702.

[14] G. Bagler, arXiv:cond-mat/0409773.

[15] K. I. Goh, B. Kahng, and D. Kim, Phys. Rev. E **72**(2005) 017103.

[16] K. I. Goh, J. D. Noh, B. Kahng, and D. Kim, Phys. Rev. E **72**(2005) 017102.

[17] Y. Gang, T. Zhou, J. Wang, Z. Fu , B. Wang, Chin. Phys. Lett., **22**(2005) 501.

[18] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**(2002) 7821-7826.

[19] P. Zhang, M. Li, J. Wu, Z. Di, Y. Fan, Physica A, **367**(2006)577-585.

[20] M. E. J. Newman, M. Girvan, Mixing patterns and community structure in networks, in Statistical Mechanics of Complex Networks, R. Pastor-Satorras, J. Rubi and A. Diaz-Guilera (eds.), pp. 66-87, Springer,Berlin(2003)

[21] J. Duch, A. Arenas, Phys Rev E. **72**(2005) 027104.

[22] Y. Fan, M. Li, P. Zhang, J. Wu, Z. Di, arXiv:physics/0607271 .

[23] L. Danon, A.t Díaz-Guilera, J. Duch and A. Arenas, J. Stat. Mech. (2005) P09008.

[24] J. Reichardt, S. Bornholdt, Phys. Rev. Lett. **93**(2004) 218701.

[25] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, Science, **297**(2002), 1551-1555 .

[26] M.E.J. Newman, M. Girvan, Phys.Rev. E **69** (2004) 026113.